

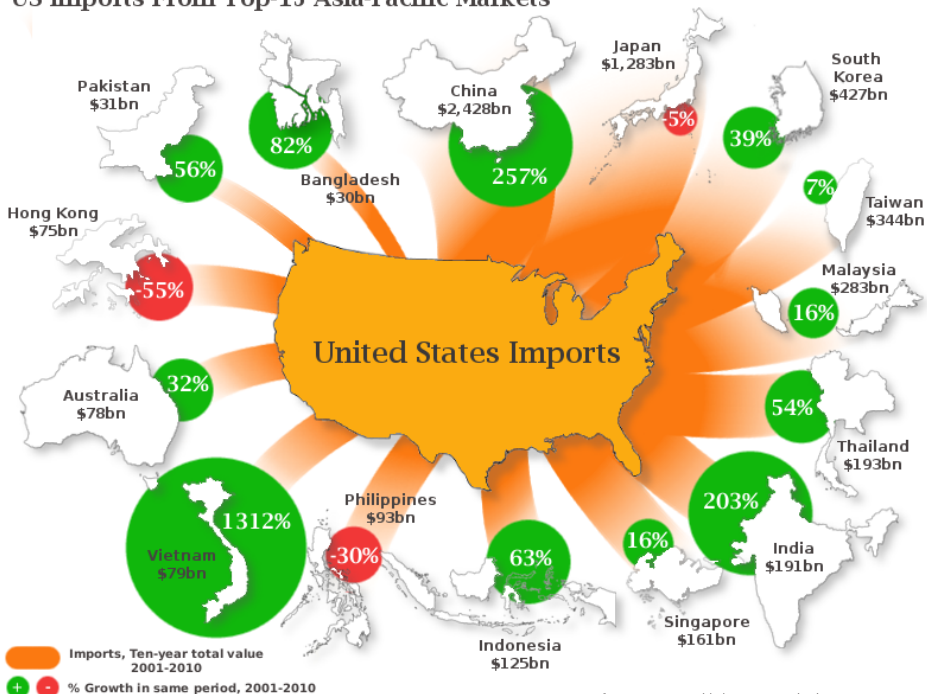


Tariff Code Finder: NLP to Detect Loopholes in the US Tariff Code

Noemi Derzsy
Data Science Fellow at Insight

Very similar goods can be subject to very different tariff rates!

US Imports From Top-15 Asia-Pacific Markets



Problem for:

- importing companies
- government
- consumers

Imported **\$2.712 trillion**
goods & services

Santa Suit



Festive Item

SOLUTION

Create an algorithm that recommends the correct tariff code based on item description.

Clothing

0% tax



8% tax

Data Sources

1. U.S. Harmonized Tariff Schedule

- Format: .pdf (**3712** pages)
- Details: the HTS contains the internationally agreed structured product nomenclature commonly known as the Harmonized System (HS)

Harmonized Tariff Schedule of the United States (2018)

Annotated for Statistical Reporting Purposes

II
6-2

| Heading/ Subheading | Stat. Suf- fix | Article Description | Unit of Quantity | Rates of Duty | | |
|------------------------|----------------------|--|------------------------|---------------|--|----------|
| | | | | 1 | | 2 |
| | | | | General | Special | |
| 0601 | | Bulbs, tubers, tuberous roots, corms, crowns and rhizomes, dormant, in growth or in flower; chicory plants and roots other than roots of heading 1212: | | | | |
| 0601.10 | | Bulbs, tubers, tuberous roots, corms, crowns and rhizomes, dormant: | | | | |
| 0601.10.15 | 00 | Tulip bulbs..... | No..... | 89.6¢/1000 | Free (A, AU, BH, CA, CL, CO, D, E, IL, JO, KR, MA, MX, OM, P, PA, PE, SG) | \$6/1000 |
| 0601.10.30 | 00 | Hyacinth bulbs..... | No..... | 38.4¢/1000 | Free (A+, AU, BH, CA, CL, CO, D, E, IL, JO, KR, MA, MX, OM, P, PA, PE, SG) | \$4/1000 |
| 0601.10.45 | 00 | Lily bulbs..... | No..... | 55.7¢/1000 | Free (A, AU, BH, CA, CL, CO, D, E, IL, JO, KR, MA, MX, OM, P, PA, PE, SG) | \$6/1000 |
| 0601.10.60 | 00 | Narcissus bulbs..... | No..... | \$1.34/1000 | Free (A, AU, BH, CA, CL, CO, D, E, | \$6/1000 |

2. U.S. cargo imports

- Format: .csv (**31** columns; **24,197,469** rows)
 - Details: information for incoming shipments regulated by U.S. Customs and Border Protection's Automated Manifest System (AMS) for 2017
- Office of Tariff Affairs and Trade Agreements -

| | identifier | trade_update_date | run_date | vessel_name | port_of_unlading | estimated_arrival_date | foreign_port_of_lading | record_status_indicator | pl |
|---|---------------|---------------------|---------------------|-----------------|----------------------------|------------------------|----------------------------|-------------------------|----|
| 0 | 2017020610912 | 2017-01-25T00:00:00 | 2017-02-06T00:00:00 | MOL MAXIM | Norfolk, Virginia | 2017-02-04T00:00:00 | Shanghai ,China (Mainland) | New | |
| 1 | 2017020611023 | 2017-01-06T00:00:00 | 2017-02-06T00:00:00 | MOL MAJESTY | Miami, Florida | 2017-02-04T00:00:00 | Shanghai ,China (Mainland) | New | |
| 2 | 2017020612148 | 2017-01-06T00:00:00 | 2017-02-06T00:00:00 | NYK ARTEMIS | Los Angeles, California | 2017-02-02T00:00:00 | Singapore,Singapore | New | |
| 3 | 2017020611030 | 2016-12-30T00:00:00 | 2017-02-06T00:00:00 | YM UTOPIA | Charleston, South Carolina | 2017-02-02T00:00:00 | Kaohsiung,China (Taiwan) | New | |
| 4 | 2017020611088 | 2017-01-20T00:00:00 | 2017-02-06T00:00:00 | OOCL WASHINGTON | Jacksonville, Florida | 2017-02-04T00:00:00 | Yantian,China (Mainland) | New | |

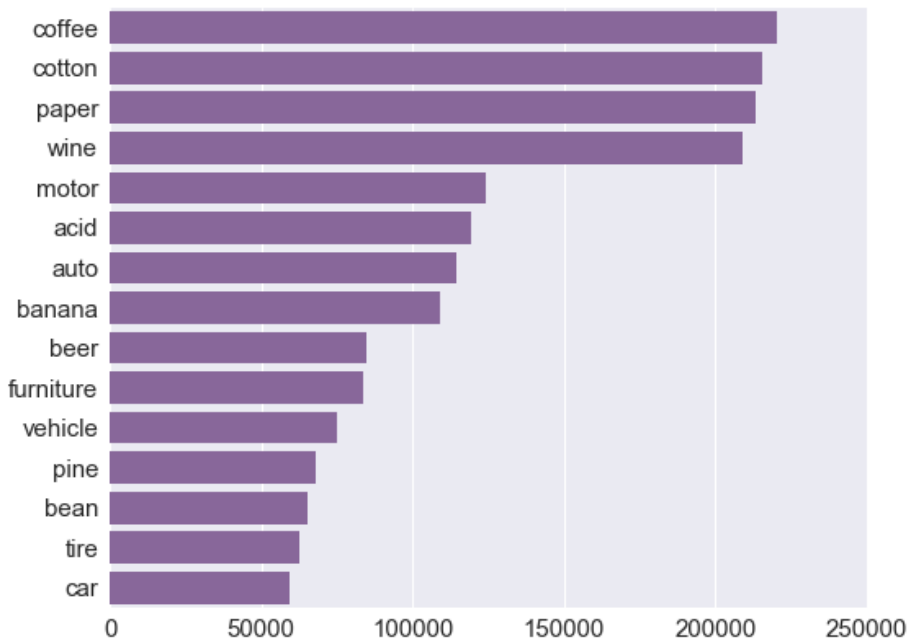
Data Challenges

- Cargo data: human supplied description, messy, full of abbreviations, misspellings -> text preprocessing (lemmatization, regular expressions)
- Cargo data: missing values (~ 80%)
- Harmonized Tariff Schedule & Cargo data: imbalanced
- Harmonized Tariff Schedule: .pdf file of over 3700 pages -> pdfx library

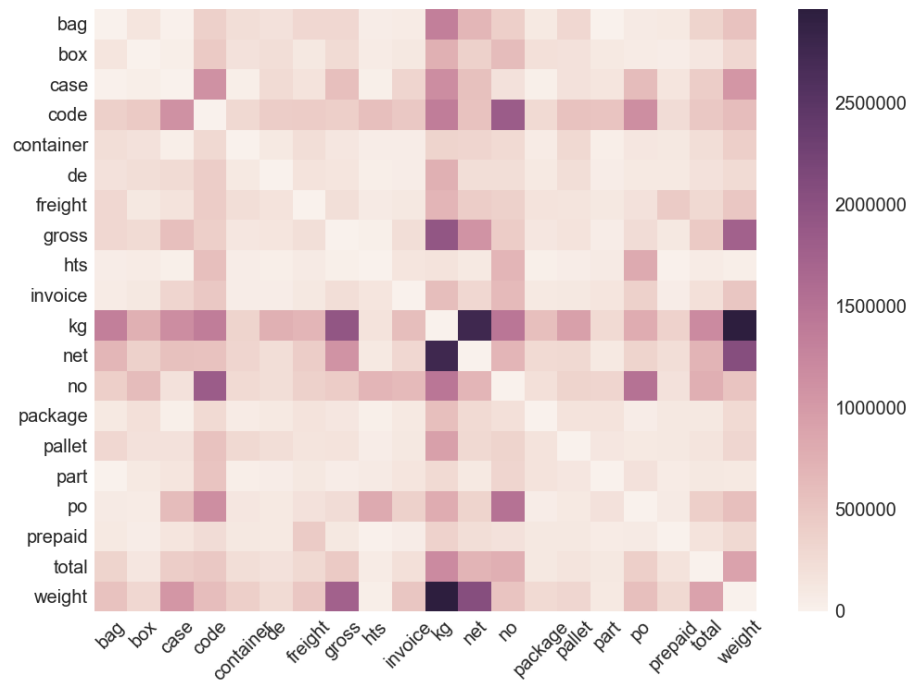
```
cargo-description-field-sample x
1 "MOTORCYCLE SPARE PARTS"
2
3 "WOMEN S KNIT, CHIFF FNT LC BK"
4
5 "P O NO. P109169897 42-7200-1048 7200 PINION L.H. P O NO. P109169653 42-7200-1005 7200 PINION R.H. P O NO. P109169330 20-1904-10041900
  PINION (BP01) 10 PLTS = 360 CTNS ""CARRIER IS NOT LIABLE FOR THE CONTENT OF PACKING & QUANTITY IN THE PALLET"" P O NO. P107010149 90207 90
  PINION 1 PLT = 16 CTNSP O NO. P109170475 68-6000-1214 PINION DC3000 DC6000 P O NO. P109171704 20-3002-1096-999 8000 PINION (20-3002-1096)
  40-1700-1005 PINION (MACHINING)""CARRIER IS NOT LIABLE FOR THE CONTENT OF PACKING & QUANTITY IN THE PALLET""P O NO. P109171704 20-7700-1434
  7700 PINION (BL11) 68-6000-1001-999 END PLUG MACHINED 68-6000-1212 PINION BUSHING"
6
7 "1032 CTN WOODCHIP BASKET __CLAUDIA ANDERSEN __M; __TEL_ 912-232-2107, EXT 214 __FAX_ 912-233-2243 __ANNA ARNSDORF,THERESA MILLER,ANNA
  OLSEN,__CLAUDIA ANDERSEN;BRANDY DAVIS N/A N/A N/A"
8
9 "995 CTN WOODCHIP BASKET ACCORDING TO SHIPPER, ""THIS SHIPMENT CONTAINS NO SOLID WOOD PACKING MATERIALS "" CARRIER HAS NOT VERIFIED THIS
  INFORMATION.PLEASE REFER TO THEAPPROPRIATE CERTIFICATION FOR DETAILS. S/C NO.US000M7X N/A"
10
11 "FROZEN STRIPED PANGASIU FILLET"
12
13 "MS METRO DESK P.O.NO.:2606177041 PLACE OF DELIVERY: SUFFOLK VIRTUAL STORAGE GLN: 0078742000008 DEPT NO:071 PO TYPE - 040TOT PKGS 948 CT 2ND
  NOTIFY: DAMCO DISTRIBUTION SERVICES INC. EAST FIRESTONE PLACE SOUTH GATE, CA 90280 USA ESCALATIONS: 323-568-2526 EMAIL: NEIL.BOOOTHBY
  DAMCO.COM NOTIFY PARTY 3: DAMCO DISTRIBUTION SERVICES INC. 5011 EAST FIRESTONE PLACESOUTH GATE, CA 90280 USA FOR ESCALATIONS: 323-568-2526
  EMAIL: DCM'LA DAMCO.COM S/C :15-0100 FREIGHT COLLECT TOTAL: 1X40 HQ CONTAINER ONLY"
14
15 "NON-NEGOTIABLE WAYBILL RELEASE PROCEDURES""SHIPPERS LOAD STOWED AND COUNT"" 1 X 40'HC CONTAINER 66 PACKAGES NEW AUTOMOTIVE PARTS HS CODE:
  87089997, 99908704 SC:10560 CY / CY ALL CHARGES PREPAID"
16 "CANNED CHUNK LIGHT SKIPJECK TUNA IN WATER POR T SIDE BRAND 48 5 OZ H.S.CODE 160414 CONNECTI NG ONTO MV EVER LEARNED V.0869W AT SINGAPORE"
17
18 "VENDOR DECLARE THIS SHIPMENT CONTAINS NO REGULATED WOOD PACKING MATERIAL. CARRIER HAS NOT VERIFIED THIS INFORMATION. PLEASE REFER TO THE
  APPROPRIATE CERTIFICATION FOR DETAILS.600 CTN 7FT MY FIRST TRAMPOLINE PO NO_4404395247, S/C_US0001EQX CY TO CY 2X40_GP FCL"
19 "MOTORCYCLE SPARE PARTS"
```

Data Exploration

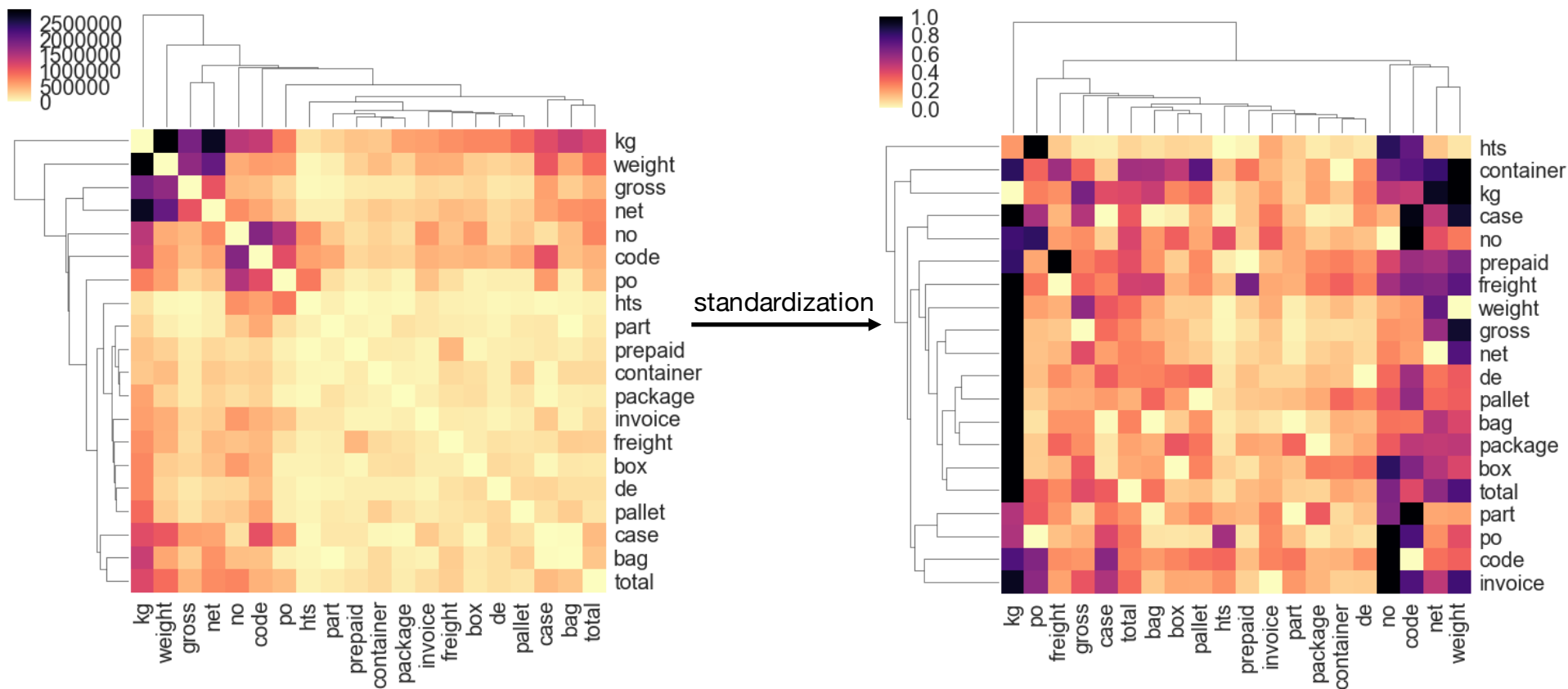
Top 15 most imported items



Word co-occurrence in cargo import descriptions



Hierarchical Clustering of Words



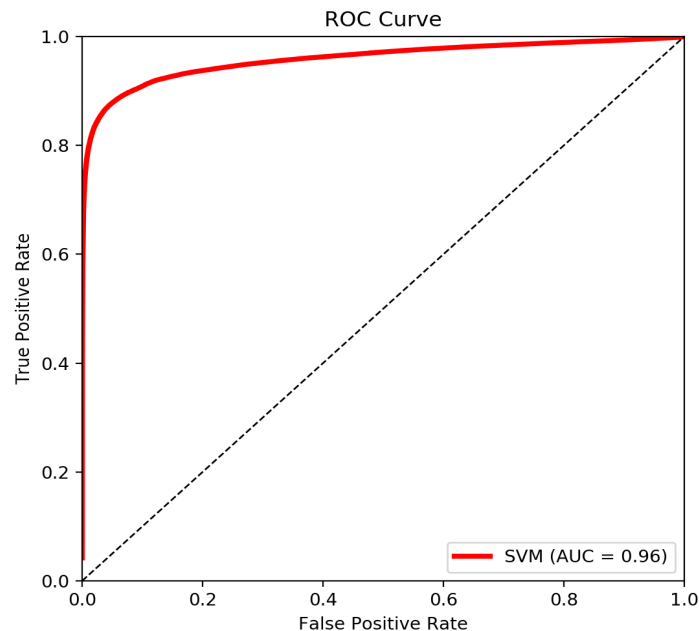
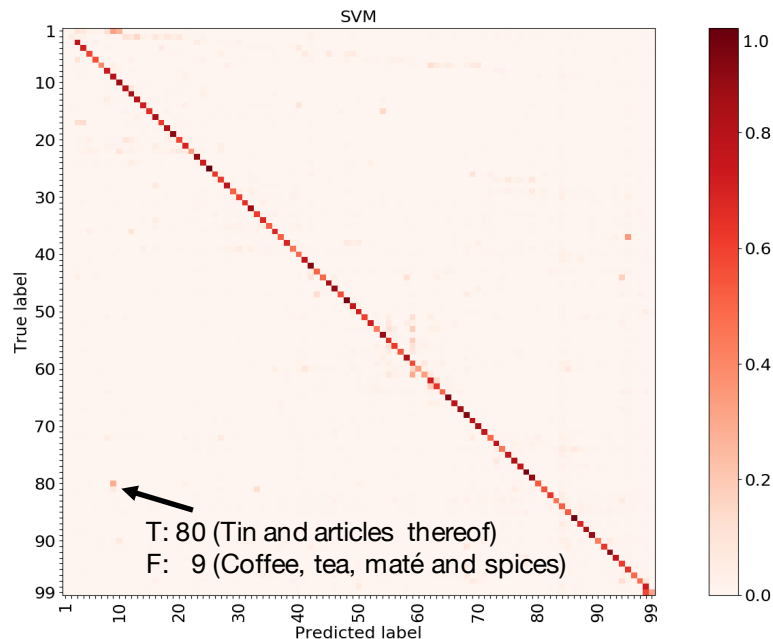
Classification Algorithm

Feature extraction:

- tf-idf
- word2vec

Classifiers:

- Naïve Bayes classifier
- k-NN classifier
- SVM



Web App - Search Engine

Tariff Code Detector

Enter the item name below.

Item Name:

Find the tariff code!

Results:

61 Apparel articles and accessories, knit or crochet

<http://item2tariffcode.site/>

Web app with Flask to input item name, and retrieve current (from Harmonized Tariff Schedule) and alternative tariff code.

Summary

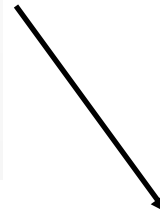
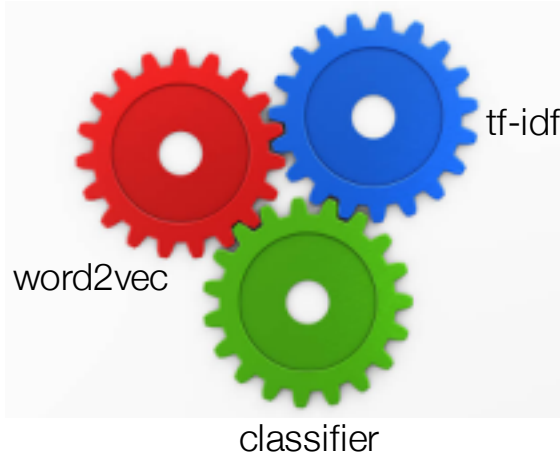
Harmonized Tariff
Schedule



Cargo imports



NLP module



Flask

web development,
one drop at a time



PostgreSQL

Noemi Derzsy

www.noemiderzsy.com

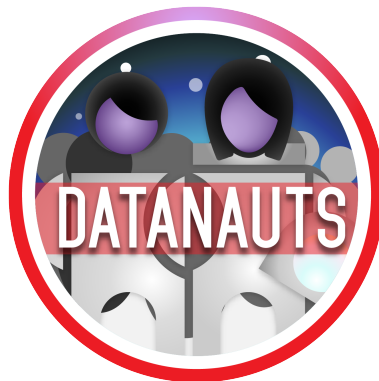
PhD in Physics, but it was really all about complex systems, especially complex networks.



Rensselaer



- Network science
- Social networks
- Data science



- Network science
- Data science
- NLP on openNASA data



INSIGHT



- Data science
- NLP on tariff codes